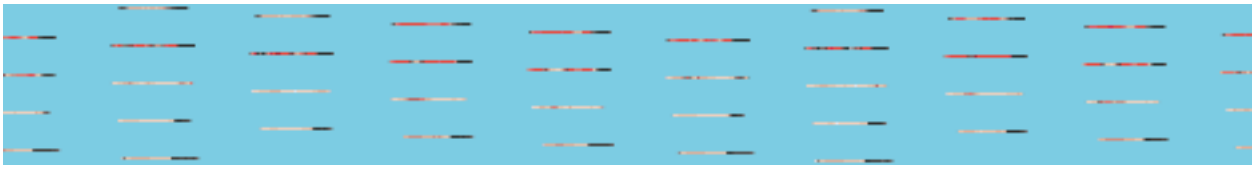
Cell type–specific analysis by single-cell profiling identifies a stable mammalian tRNA–mRNA interface and increased translation efficiency in neurons

William Gao, Carlos J. Gallardo-Dodd and Claudia Kutter

Genome Res. 2022 32: 97-110 originally published online December 2, 2021
Access the most recent version at doi:10.1101/gr.275944.121



Cell type–specific analysis by single-cell profiling identifies a stable mammalian tRNA–mRNA interface and increased translation efficiency in neurons

William Gao, Carlos J. Gallardo-Dodd, and Claudia Kutter

Department of Microbiology, Tumor, and Cell Biology, Karolinska Institute, Science for Life Laboratory, 171 77, Stockholm, Sweden

The correlation between codon and anticodon pools influences the efficiency of translation, but whether differences exist in these pools across individual cells is unknown. We determined that codon usage and amino acid demand are highly stable across different cell types using available mouse and human single-cell RNA-sequencing atlases. After showing the robustness of ATAC-sequencing measurements for the analysis of tRNA gene usage, we quantified anticodon usage and amino acid supply in both mouse and human single-cell ATAC-seq atlases. We found that tRNA gene usage is overall coordinated across cell types, except in neurons, which clustered separately from other cell types. Integration of these data sets revealed a strong and statistically significant correlation between amino acid supply and demand across almost all cell types. Neurons have an enhanced translation efficiency over other cell types, driven by an increased supply of $tRNA^{Ala}$ (AGC) anticodons. This results in faster decoding of the Ala-GCC codon, as determined by cell type–specific ribosome profiling, suggesting that the reduction of $tRNA^{Ala}$ (AGC) anticodon pools may be implicated in neurological pathologies. This study, the first such examination of codon usage, anticodon usage, and translation efficiency resolved at the cell-type level with single-cell information, identifies a conserved landscape of translation elongation across mammalian cellular diversity and evolution.

[Supplemental material is available for this article.]

During translation elongation, the ribosome moves three nucleotides at a time along the mRNA transcript, as each codon complementarily binds to a corresponding anticodon triplet on a tRNA, which is charged with a specific amino acid (AA) that is then added to the growing polypeptide chain (Dever et al. 2018). Because translation elongation relies on the codon–anticodon interaction, matching or mismatching of codon and anticodon pools may influence the elongation rate. Indeed, it has been shown in bacteria such as *Escherichia coli* and unicellular eukaryotes such as *Saccharomyces cerevisiae* that matching of these pools results in increased translation efficiency, leading to higher production of proteins. Conversely, mismatching of these pools yields lower protein production (Quax et al. 2015).

Codon and anticodon pools have also been measured in multicellular organisms, such as mammals, which pose several additional challenges. In particular, the mRNA and tRNA levels may differ across tissues and cell types, and quantitation can be biased because of experimental challenges (Wong et al. 2012). The development of bulk high-throughput sequencing (Stark et al. 2019) in the past two decades has allowed for examination of these mRNA and tRNA pools across a few mammalian tissues, revealing some tissue-specific differences but an overall stability in both codon and anticodon usage (Dittmar et al. 2006; Schmitt et al. 2014; Rak et al. 2018; Pinkard et al. 2020). In these studies, codon pools were measured primarily using bulk RNA sequencing. However, because of several special features of tRNA genes (isodecoders), anticodon pools have been measured with a variety of RNA- and DNA-based methods that are subject to different biases. For example, RNA-based methods for tRNA gene expression include microarray (Chou et al. 2004; Dittmar et al. 2006) and high-throughput sequencing approaches (Zheng et al. 2015; Gogakos et al. 2017; Shigematsu et al. 2017; Xu et al. 2019; Kugelberg et al. 2021). These methods must overcome the challenge of producing complementary DNA from tRNAs, which are not only highly structured but also extensively modified (Rak et al. 2018; Suzuki 2021). As a result, no "gold-standard" approach exists to measure tRNA abundances, with a recent study showing that the tRNA levels obtained from different methods are lowly correlated ($\rho \sim 0.22$–0.62) (Pinkard et al. 2020). Alternatively, chromatin immunoprecipitation with massively parallel DNA sequencing (ChIP-seq) can be performed in bulk with an antibody targeting an active subunit of RNA polymerase III (Pol III) that transcribes tRNA genes (Dieci et al. 2007). Examining active Pol III occupancy on DNA quantifies the amount of tRNA transcription and is not prone to biases resulting from tRNA structure and modification (White 2011). Moreover, because ChIP-seq reads often extend past the exact sequence of the tRNA gene locus, flanking sequence information can be used to resolve tRNA genes with otherwise identical sequences. However, Pol III ChIP-seq only probes pre-tRNA transcription. Pre-tRNAs undergo several more steps before becoming ready-to-translate tRNAs (Wolin and Matera 1999; Phizicky and Hopper 2010).

Using these methods of quantifying codon and anticodon usage, previous studies have examined how correlated these pools are, using this as a proxy for the efficiency of translation elongation (often referred to as "translation efficiency"). Analyses of codon and anticodon pools across the tree of life have revealed several unifying features, such as the necessity of wobble base-pairing between codons and anticodons. Indeed, all organisms use

Corresponding author: claudia.kutter@ki.se
Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.275944.121.

fewer than 61 anticodons to decode the 61 sense codons because decoding can occur with wobble interactions between the first anticodon position (tRNA nucleotide 34) and the third codon position (Rak et al. 2018). In particular, the most well characterized modifications include G:U and adenosine-to-inosine 34 (A34-to-I) wobbling. The latter, performed by adenosine deaminases (ADATs), expands the repertoire of ANN anticodons to decode not only NNU codons but also NNC and NNA codons (Torres et al. 2014).

In unicellular organisms such as *E. coli* and *S. cerevisiae*, the most frequent codons also correspond to the most abundant anticodons, and this explains a great deal of patterns in synonymous codon usage bias (Rocha 2004). In contrast, work in mammals has suggested that adaptation to anticodon pools cannot explain the majority of synonymous codon usage, which appear to be more strongly influenced by mutational biases and drift (dos Reis et al. 2004), in particular GC-biased gene conversion (Pouyet et al. 2017). Despite weak correlations between synonymous codon usage bias and anticodon levels (Novoa and Ribas de Pouplana 2012), the overall correlation between codon and anticodon pools in mammals is nonetheless quite strong and stable (Rudolph et al. 2016). Yet, all of these observations in mammals have been made using bulk sequencing.

However, a major limitation of using bulk methods to quantify codon and anticodon pools is that by aggregating data from multiple different cell types within a tissue, they may blur out heterogeneity across cell types. Because cell types use not only different levels of the same proteins but also different types of proteins to perform their various functions, it is possible that their codon pools are different. Additionally, the tRNA gene usage across cell types may also differ to match their codon pools (Dittmar et al. 2006; Rak et al. 2018).

The advent of multiomic single-cell technologies and the production of publicly available whole-organism single-cell atlases allow us to examine codon and anticodon usage at cell type specificity and uncover heterogeneity that cannot be detected in bulk. In this study, we have leveraged this power using both adult mouse and fetal human, hereafter referred to as mouse and human, single-cell RNA-sequencing (scRNA-seq) atlases (The Tabula Muris Consortium 2018; Cao et al. 2020), as well as complementary single-cell assay for transposase accessible chromatin (scATAC-seq) atlases (Cusanovich et al. 2018; Domcke et al. 2020). We examined the reliability of using information on chromatin accessibility as a proxy for quantification of anticodon usages, and investigate the interdependencies of matched codon and anticodon pools at the cell-type level. Finally, we calculate translation efficiency by correlating AA demand and supply from the mRNA and tRNA side, respectively. We consider this to be an estimate of the maximum translation rate that can be achieved when tRNA processing steps have a negligible influence on the levels of mature charged tRNAs and, hence, herein refer to this measure as theoretical translation efficiency (tTE). By examining these important contributors to translation elongation for the first time systematically at the cell-type level, this study paves the way for future high-resolution studies of mammalian translation.

Results

Codon usage is highly similar across cell types

To investigate whether codon usage differs across mammalian cell types, we analyzed data from both adult mouse and fetal human scRNA-seq atlases. These data sets consist of cells from 20 and 15 tissues, respectively, where cell type annotations have been provided for all individual cells that passed quality filters (Fig. 1A,B; Supplemental Fig. S1A; Supplemental Table S1).

Different cell types express a diverse assembly of genes at dissimilar levels, and the high expression of specific marker genes is often used for cell type annotation. When pooling the gene expression profiles from individual cells of the same cell type, we observed the expected clustering of cell types based on function and tissue origin (Fig. 2A; Supplemental Tables S2, S9). For example, endothelial and stromal cells from several organs clustered together, whereas in other cases, cell types from the same tissue congregated owing to the expression of tissue-specific genes, as observed previously (The Tabula Muris Consortium 2018). Clustering of cell types by origin and function was less defined in the human data set, perhaps owing to the collection of these samples in early development (72–129 d) (Cao et al. 2020).

To calculate the codon usage across cell types, we weighted the 61 sense codon frequencies of each protein-coding gene by their cell type–specific expression (Fig. 1B, top; Supplemental Tables S3–S4, S10–S11). The sense codon frequencies were determined from the trinucleotide frequencies of the coding sequences (CDSs) of each protein-coding gene. The longest open reading frame (ORF) per protein-coding gene was used to determine its codon frequencies, as isoform information is not available for all cell types in the scRNA-seq data sets but remains an active area of research (Uhlén et al. 2015).

We found, in contrast to gene expression, a strong similarity in codon usage across cell types. The Euclidean distances between codon usages across almost all cell types are small (Fig. 2B), and a principal component analysis (PCA) plot of codon usages across cell types reveals that nearly all of them colocalize without distinct clustering (Fig. 2C). A previous analysis of codon usage in bulk had also suggested highly stable codon pools across tissues (Schmitt et al. 2014), but it remained unknown whether particular cell types had distinct codon usage profiles that were undetected in bulk. Here, we found that codon usage was highly similar across most cell types. Moreover, the mean codon usage across all cell types was highly similar to the exonic background (the overall genome codon frequencies unscaled by gene expression) (Fig. 2D), as established in bulk (Schmitt et al. 2014). In other words, our analysis at the cell-type level largely recapitulated findings from bulk analysis owing to true homogeneity in codon usage across cell types, rather than averaging of heterogeneous codon usages.

Cell types that deviate from codon usage have low codon pool complexity

Although most cell types have similar codon usages, some cell types were distinct. In both mouse and human, cardiac muscle cells and pancreatic acinar cells had larger Euclidean distances from other cell types and resided away from the main cluster in PCA (Fig. 2B,C). Stromal cells from some human tissues also segregated from the main cluster, whereas others did not (Fig. 2C). No mouse stromal cell populations were outside the main cluster. Thus, the differential codon usage in some fetal stromal cell types may be present only in early development. Skeletal muscle cells, which were not present in the mouse data set, clustered in codon usage with cardiac muscle cells in the human data set (Fig. 2C).

Because codon usage was computed by weighting each gene's codon frequencies by its expression, differential codon usage in cardiac/skeletal muscle cells and pancreatic acinar cells could be

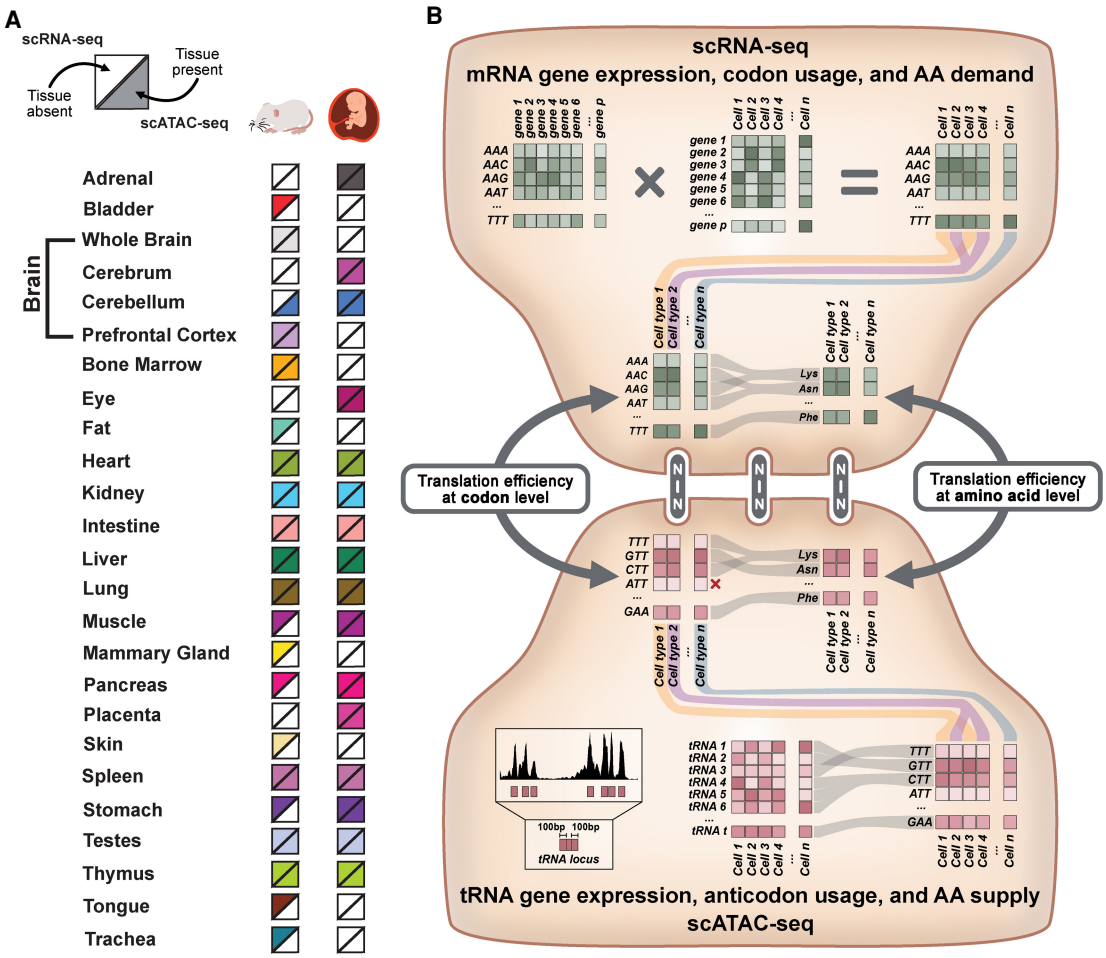


Figure 1. Overview of the approach for single-cell analysis of translation efficiency. (*A*) To examine a diverse set of cell types across multiple tissues, scRNA-seq and scATAC-seq atlases produced for mouse and human were analyzed. For both organisms, each square is color-filled if scRNA-seq (*top* half) or scATAC-seq data (*bottom* half) are present. (*B*, *top*) mRNA gene expression, codon usage, and amino acid (AA) demand can be quantified using single-cell RNA sequencing (scRNA-seq). Each gene's sense codon frequencies are weighted by gene expression counts. Codon usage from individual cells is pooled at the cell-type level and can be combined based on AA demand. (*Bottom*) tRNA gene expression, anticodon usage, and AA supply can be quantified using the single-cell assay for transposase accessible chromatin (scATAC-seq). First, the Tn5 transposase insertions mapping within tRNA gene loci and their surrounding 100-bp flanking regions are quantified, creating a tRNA gene expression matrix. Cells are pooled at the cell-type level and can be combined into anticodon isoacceptor families and AA isotypes. (*Center*) scRNA-seq and scATAC-seq data can be integrated to calculate a tTE score. For more detailed description, see Methods section.

influenced by the diversity of protein-coding genes that contributes to their codon pools. A cell type could be an outlier in codon usage because of widespread differences in codon usage across numerous expressed genes or because its codon usage is skewed by a few highly expressed genes with large codon demands. To determine which was the case, we calculated each gene's codon pool contribution per cell type, defined as the number of codons per gene multiplied by its cell type–specific expression level. Thus, longer genes that are also highly expressed will have the largest codon pool contributions. We defined codon pool complexity as the proportion of the codon pool that is not contributed by the top N genes, ranked by codon pool contribution. Thus, a low codon pool complexity means that the top N genes' codon pool contributions are a large percentage of the total codon pool, whereas a high codon complexity indicates that many genes contribute a small amount to the codon pool.

We found that cell types with low codon pool complexity (calculated for several values of N ranging from five to 100) tend to be those that are outliers in codon usage, with the lowest correlations in codon usage to the mean codon usage across all cell types (Fig. 2E). By ignoring the codon contribution of the top 10 genes, their codon usages become indistinguishable from the other cell types (Fig. 2F). Therefore, these cell types were outliers in codon usage because relatively few protein-coding genes with skewed codon frequencies comprised the bulk of their codon pools. For example, cardiac and skeletal muscle cells express high levels of *TTN*, the gene encoding titin. It is the largest known protein (27,000 to 35,000 AAs, depending on the isoform) and important for the passive elasticity of muscle (Lewinter and Granzier 2010). Because of its remarkable length and high expression in cardiac and skeletal muscle cells, this gene accounts for between 48% and 71% of their codon pools (56% in mouse cardiac muscle cells, 48%–71% in human cardiac/skeletal muscle cells). Thus, when titin's usage of codons differs from the mean codon usage across cell types, cardiac and skeletal muscle cells become outliers for those codons (Fig. 2G).

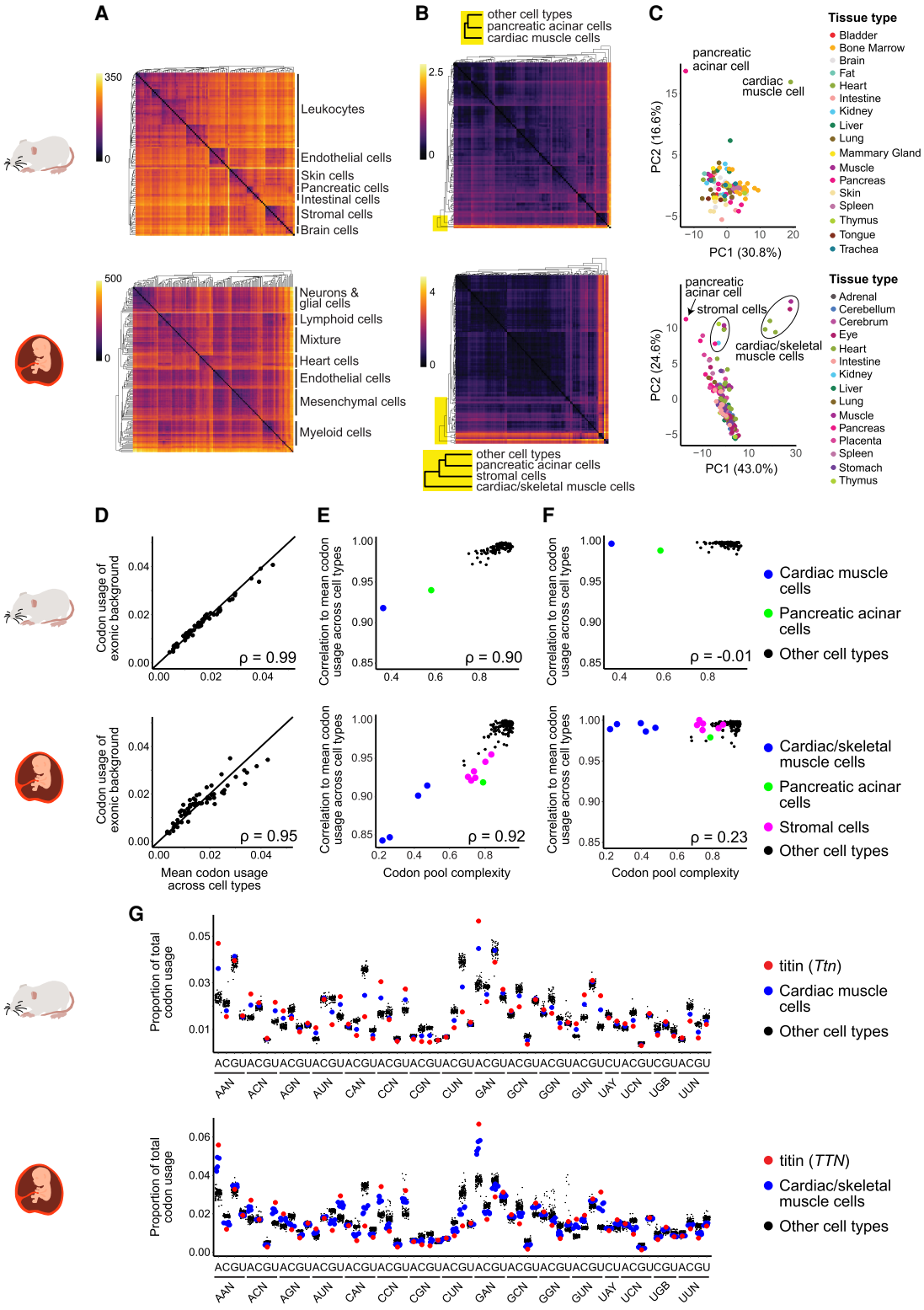


Figure 2. Codon usage is highly stable across cell types, with outliers driven by reduced codon pool complexity. (*A*,*B*) Heatmaps show Euclidean distance in gene expression (*A*) and codon usage (*B*) across cell types for mouse (*top*) and human (*bottom*). (*C*) Principal component analysis (PCA) plots cluster codon usage across cell types. (*D*–*F*) Scatter plots: (*D*) correlation of mean codon usage across all cell types to the exonic background (unweighted genome-wide codon usage; each point corresponds to one of the 61 sense codons); (*E*) each cell type's codon pool complexity (percentage of the total codon pool contributed by the top 10 genes) to the correlation of its cell type to the mean codon usage across all cell types; and (*F*) same quantities as *E*, but with each cell type's codon usage calculated while ignoring the codon contribution of the top 10 codon pool contributing genes. (*G*) Jitter plots show the proportion of total codon usage for each of the 61 sense codons in each cell type. Titin (red) drives the outlier status of cardiac and skeletal muscle cells (blue). IUPAC nucleotide codes are used in the *x*-axis (N = A, C, G, U; B = C, G, U; Y = C, U).

scATAC-seq is robust for measuring tRNA gene usage

Having established that codon usage is highly stable across most cell types, we next investigated whether anticodon usage would also be similar. Previous work has examined tRNA gene usage in different cell lines and in tissues, identifying tissue-specific differences (Dittmar et al. 2006; Schmitt et al. 2014; Rudolph et al. 2016; Pinkard et al. 2020). Although existing RNA- and DNA-based tRNA quantification methods have their respective advantages and disadvantages, none have been adapted for analysis in single cells. To address this major limitation, we investigated whether the assay for transposase-accessible chromatin (ATAC-seq) could be used to examine tRNA gene usage. This assay, uses a hyperactive Tn5 transposase that cuts at open chromatin (Buenrostro et al. 2015). Thus, loci enriched for Tn5 insertions correspond to regions of open chromatin, at which transcription factors and RNA Pol can bind. Indeed, chromatin accessibility in gene bodies often correlated well with transcription levels (Klemm et al. 2019). Furthermore, several different methods have been developed for single-cell analysis (Pott and Lieb 2015; Baek and Lee 2020).

Because scATAC-seq has not been used before for analyzing tRNA genes, we assessed its robustness by determining whether it is internally consistent (different scATAC-seq pipelines produce similar quantifications of tRNA accessibility), it is externally consistent (concordant with expectation), and it is of adequate resolution (can distinguish between tRNA genes located in tight genomic clusters). We found that tRNA gene accessibility from three available scATAC-seq data sets of mouse brain (aggregated in pseudobulk) (Cusanovich et al. 2018; Lareau et al. 2019) was strongly correlated at the tRNA gene, anticodon isoacceptor, and AA isotype levels (Supplemental Figs. S1, S2; Supplemental Tables S5–S7, S12–S14), showing the internal consistency of ATAC-seq for tRNA gene quantification. These data were also highly correlated at all three levels with a bulk mouse brain ATAC-seq data set (Liu et al. 2019) and, importantly, with tRNA abundances measured by Pol III ChIP-seq, the established DNA-based tRNA quantification method. This suggests that the chromatin accessibility of tRNA genes is a suitable measure for their expression (Fig. 3A). The correlations across these different data sets were consistently high (Spearman's rank correlation coefficients, $\rho \sim 0.7-0.9$), which is much higher than those obtained when comparing across different RNA-based quantification methods (Pearson correlation coefficients, $r \sim 0.2-0.6$) (Pinkard et al. 2020). To validate ATAC-seq irrespective of other approaches, we also examined whether highly confident functional tRNA genes were more accessible than low-confidence tRNA gene predictions (Lowe and Chan 2016). We found that high-confidence tRNA genes were generally much more accessible than low-confidence ones (Fig. 3B). Finally, because tRNA genes tend to be genomically organized in dense clusters, we assessed whether scATAC-seq data provided enough resolution to differentiate accessibility between very close tRNA genes. Indeed, scATAC-seq displays individual peaks for tRNA genes in clusters in which tRNA genes were <300 bp apart (Fig. 3C).

Although this shows the robustness of scATAC-seq for quantifying tRNA gene usage, a practical concern was that scATAC-seq data could be very sparse because there are normally only two copies of DNA per cell (Baek and Lee 2020). Moreover, because tRNA genes are very short and occupy a small percentage of mammalian genomes, the number of cuts to tRNA genes per cell could be low at normal sequencing depths (Supplemental Fig. S1B). To determine how many cuts to tRNA genes (gene body plus 100 bp upstream and downstream, as typically performed for Pol III ChIP-seq data) (Kutter et al. 2011) are required to obtain reliable measurements, we took different sample sizes of cuts from pseudobulked scATAC-seq data and measured the correlation of these samples to the overall tRNA gene usage (pseudobulked scATAC-seq data without down-sampling). At small sample sizes, the correlations at the tRNA gene, anticodon, and AA level were low and highly variable, whereas a sample size of greater than 5000 cuts reliably yielded high correlations ($\rho > 0.95$) at all three levels (Fig. 3D). Thus, we analyzed cell type–specific tRNA gene usage by pooling all cells annotated as belonging to a particular cell type. Only cell types with more than 5000 total cuts after pooling were considered reliable for tRNA gene expression analysis.

Neuronal cell types are outliers from an otherwise stable anticodon pool

To examine tRNA gene expression, anticodon usage, and AA supply across many different cell types, we analyzed scATAC-seq atlases that corresponded to the scRNA-seq atlases (Fig. 1; Supplemental Fig. S1). The mouse and human scATAC-seq atlases contained 13 and 15 tissues, respectively (Cusanovich et al. 2018; Domcke et al. 2020). For the cell types remaining after filtering for sufficient scATAC-seq cuts, we combined the expressions of tRNA genes on the anticodon level and computed their usage across cell types (Fig. 1B, bottom). We noticed that although anticodon usage was generally more variable than codon usage, it was still similar across cell types, as indicated by similar Euclidean distances in anticodon usage across most cell types (Fig. 4A) and the presence of a large PCA cluster corresponding to most cell types (Fig. 4B). However, we also observed a distinct neuron-specific cluster in the mouse. This brain neuron cluster was also present in humans, although less pronounced, again possibly owing to decreased cell differentiation in early development. Additionally, the human data set also included other neuronal cell types from the eye and enteric nervous system as well as neuroendocrine cells absent from the mouse data set.

To examine what caused brain neurons to cluster separately from other cell types, we performed differential analysis at the anticodon level. We found that in both the mouse and human, the anticodon AGC charged with alanine, $tRNA^{Ala}$ (AGC), is enriched by >25% in brain neurons compared against all other cell types (Supplemental Fig. S3). In the mouse brain, the two other alanine anticodon isoacceptors, $tRNA^{Ala}$ (UCG) and $tRNA^{Ala}$ (CCG), were also enriched in glial cells (Fig. 4C). The fourth alanine anticodon isoacceptor, $tRNA^{Ala}$ (GCG), is absent in either mammalian genome. Accordingly, when we quantified on the AA isotype level, we confirmed $tRNA^{Ala}$ enrichment in neurons (Fig. 4D). Although the tRNA anticodon pools of all cell types was fairly stable, the slight deviation of one AA isotype could have significant effects on global translation (Kapur et al. 2017).

To verify that Ala isotype enrichment was observed in other data sets, we also analyzed available bulk data sets. We found that Ala supply is enriched in the embryonic brain in a bulk ATAC-seq data set of mouse embryonic development (Supplemental Fig. S4; Gorkin et al. 2020), a bulk ChIP-seq data set of mouse brain and liver from early development to adulthood (Supplemental Fig. S5; Schmitt et al. 2014), and a RNA-based multitissue QuantM-tRNA-seq data set of adult mouse (Supplemental Fig. S6; Pinkard et al. 2020). This, in addition to the comparison to Pol III ChIP-seq described above (Fig. 3A), underscores the concordance of scATAC-seq-based tRNA quantification to other methods.

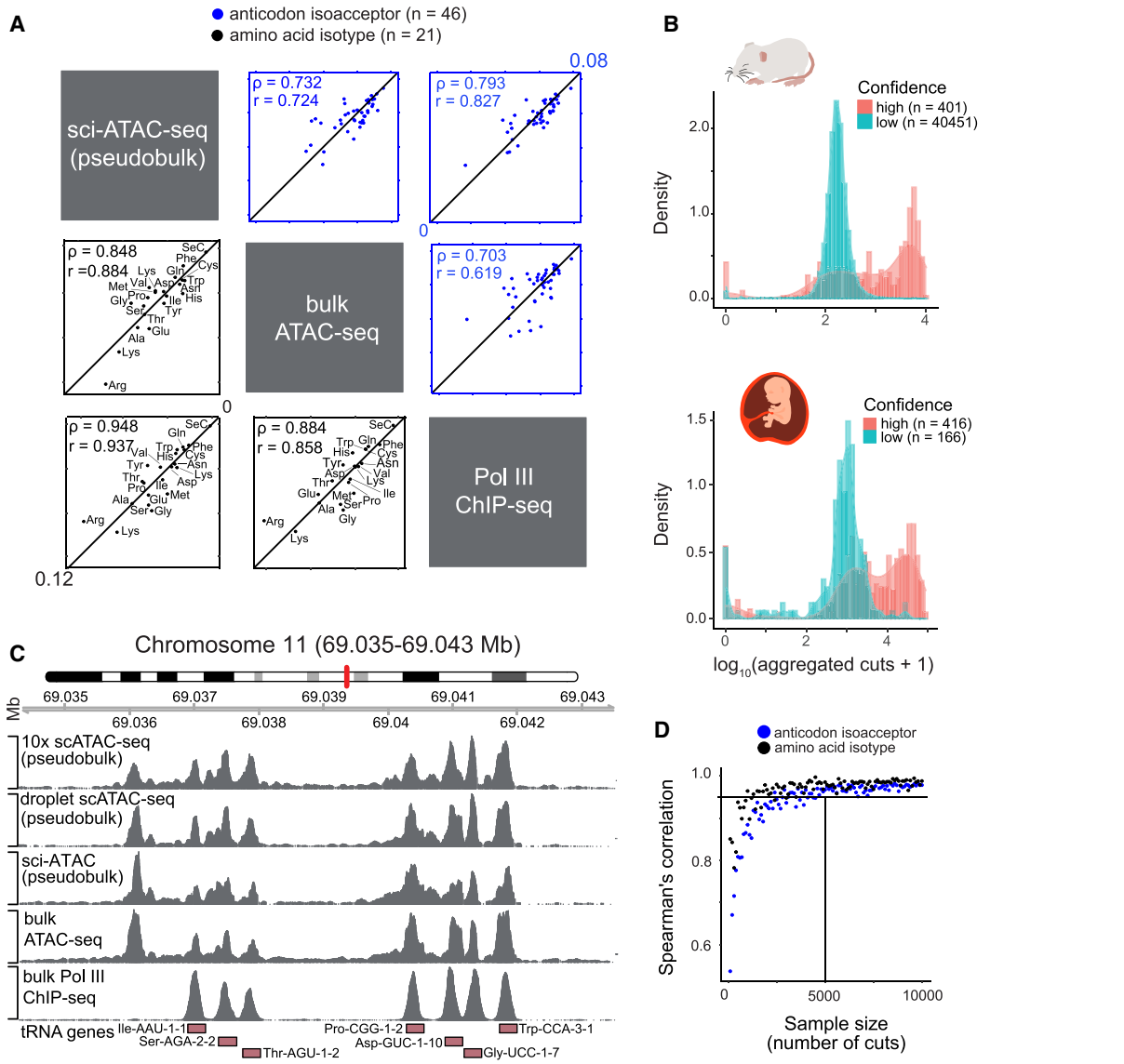


Figure 3. Single-cell ATAC-sequencing is robust for measuring tRNA gene usage. (*A*) Scatter plots correlate mouse brain scATAC-seq (aggregated in pseudobulk), bulk ATAC-seq, and Pol III ChIP-seq data sets on the anticodon isoacceptor (blue) and AA isotype (black) level. Spearman's rank (ρ) and Pearson (r) correlation coefficients are indicated on the *top left* corners. (*B*) Density plots (in log-scale) show the total number of cuts from aggregated scATAC-seq data from the mouse and human scATAC-seq atlases, based on confidence predicted from tRNAscan-SE. (*C*) Genome browser view illustrates a mouse tRNA gene cluster on Chromosome 11. The location of tRNA genes is shown at the *bottom*, including upstream and downstream 100-bp flanking regions. Different tRNA genes, including those as close as 220 bp apart (*Asp-GUC-1-10* and *Gly-UCC-1-7*), have distinct peaks in three scATAC-seq data sets (pseudobulk), a bulk ATAC-seq data set, and a Pol ChIP-seq data set (all mouse brain). The *leftmost* peak present in all ATAC-seq data sets but absent in the Pol III ChIP-seq data set corresponds to the promoter of protein-coding gene *Ctc1*. This peak is ignored when quantifying tRNA gene usage because it falls outside of 100-bp flanking region of a tRNA gene. (*D*) Scatter plot determines how many total cuts are needed after pooling cells of the same cell type to obtain reliable information of anticodon isoacceptor and AA isotype usage. A sample size of 5000 cuts per cell type consistently yielded a Spearman's rank correlation coefficient >0.95 to the aggregated scATAC-seq data from mouse and human.

Several tRNA anticodons are enriched in neurons

Although tRNA genes encoding for the same anticodon and AA isotype are often considered functionally redundant, previous work has shown that the expression of individual tRNA genes can vary drastically across tissues (Schmitt et al. 2014; Pinkard et al. 2020). In contrast, anticodon isoacceptor usage and AA supply are much more similar across tissues, suggesting an unknown buffering mechanism that coordinates tRNA gene expression to yield a stable supply of anticodons. This apparent buffering was also observed at the cell-type level in our analysis (Fig. 4C,D). However, the assumption that tRNA genes of the same anticodon isoacceptor family are functionally identical may be untrue (Pan 2018).

Therefore, we examined differential expression at the individual tRNA gene level. As with anticodon usage, PCA at the tRNA gene level for both data sets indicated a brain neuron cluster (Supplemental Fig. S7). By performing a differential gene analysis comparing brain neurons against all other cells in the mouse and human data sets, we observed that several tRNA genes that

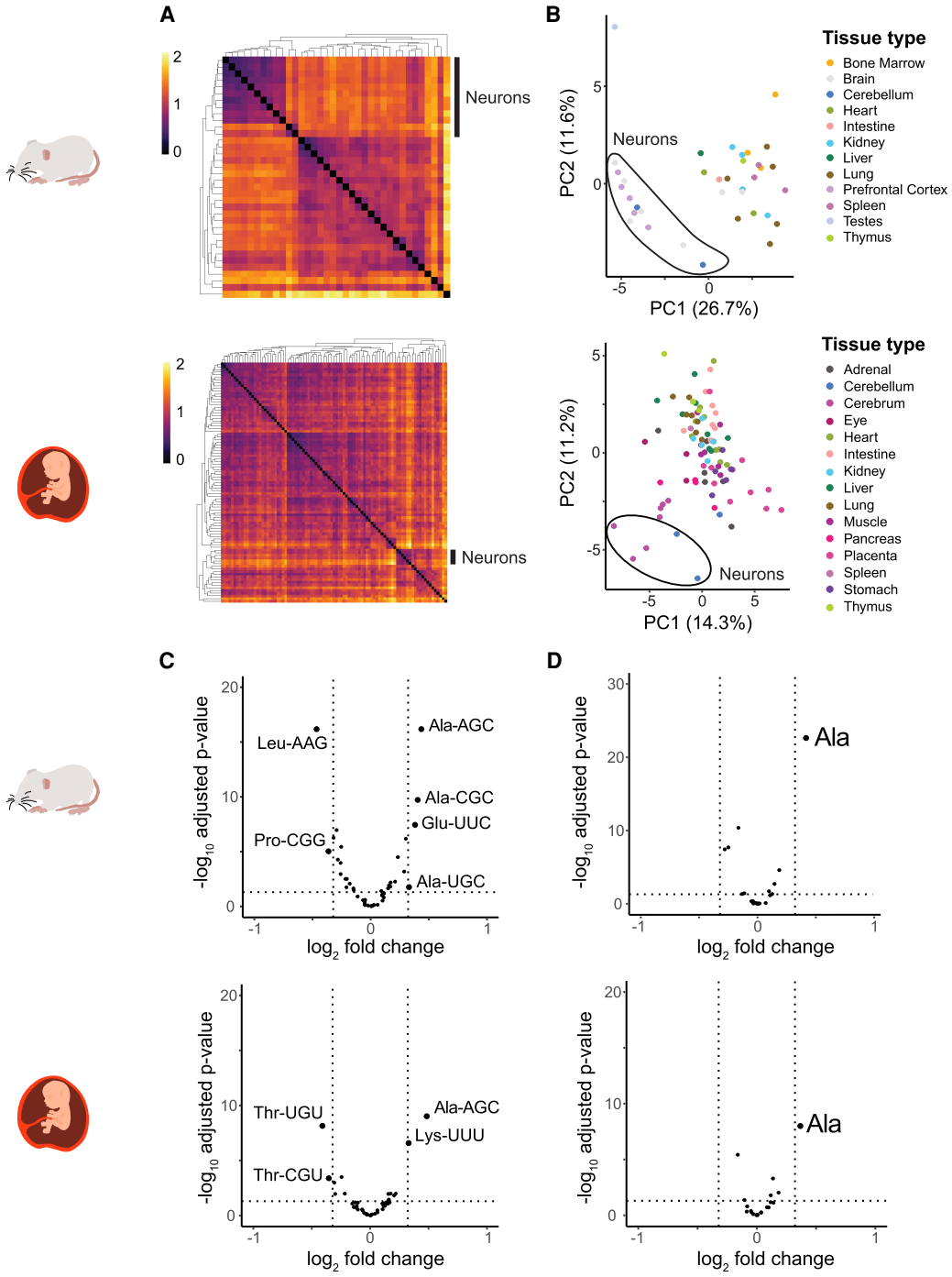


Figure 4. Anticodon usage is also similar across cell types, with brain neurons clustering separately. (*A*) Heatmaps show the Euclidean distance between anticodon usage across cell types. Only cell types with more than 5000 scATAC-seq cuts were analyzed (see Fig. 3D). The neuronal cluster is indicated with a black line on the *right* side of the heatmap. (*B*) PCA plots separate anticodon usage across cell types. The brain neuron cluster is indicated. (*C*,*D*) Volcano plots display differences in anticodon usage (*C*) and AA supply (*D*) between brain neurons and all other cell types ($-\log_{10}$ adjusted *P*-values and $\log_2$ fold change [FC] as determined using DESeq2). Vertical lines indicate a FC >25%. All panels: mouse (*top*), human (*bottom*).

are enriched in neurons not only have the same anticodon but also are syntenic in the human and mouse genomes (Fig. 5A,B). These syntenic, neuronally enriched tRNA genes include *Arg-UCU-4-1* (*n-Tr20*), which is one of the few tRNA genes that has been well characterized. A tRNA maturation-inhibiting mutation in *Arg-UCU-4-1* combined with deletion of a ribosome recycling factor causes ribosome stalling at cognate Arg-AGA codons and leads to ataxia and early death in mice (Ishimura et al. 2014). Further work has shown that loss of this tRNA reduces seizure susceptibility (Kapur et al. 2020). We found two other syntenic, neuronally enriched tRNA genes, *Ala-AGC-3-1* (same name in both GRCm38 and GRCh37) and *Ile-UAU-2-1* (in GRCm38) corresponding to

Ile-UAU-2-3 (in GRCh37). Because of the cell type resolution of scATAC-seq, we observed a clear difference in expression of these tRNA genes in neurons compared with glial cells. Additionally, in the human data set, which contained cells for the stomach and adrenal gland that are absent from the mouse data set, we found high expression of neuronally enriched tRNA genes in non-brain neurons, including neurons in the eye (ganglion cells) and the enteric nervous system, as well as in neuroendocrine cells (chromaffin cells, sympathoblasts, Schwann cells, and islet endocrine cells) (Fig. 5B). In summary, a specific subset of tRNA genes is uniquely transcribed in cells of the neuronal lineage.

AA supply and demand are highly correlated across cell types

Having analyzed codon and anticodon usage separately, we next sought to correlate them to assess the potential impact of codon and anticodon pools on translation elongation. Because there is no one-to-one correspondence between codons and anticodons, methods such as the tRNA adaptation index (tAI) have been developed to match multiple cognate anticodons to their respective codons by estimating the stability of various codon–anticodon interactions. With the assumption that highly expressed genes should be better adapted to the anticodon pool, wobbling coefficients can be calculated to measure the correlation between an organism's codon and anticodon pools (dos Reis et al. 2004). However, synonymous codon usage biases in mammals are not believed to be under strong translational selection from tRNA pools but rather from mutational biases (Dos Reis and Wernisch 2009; Pouyet et al. 2017). Moreover, we wanted to assess differences in translation efficiency across cell types rather than assume optimal wobble coefficients for translation of the most highly expressed protein-coding genes per cell type. Therefore, we instead calculated tTE scores defined as the Spearman's rank correlation between AA supply and demand from the tRNA and the mRNA side, respectively. Although this is imperfect as not all tRNAs charged with a particular AA can decode every codon demanding that same AA (Watanabe and Yokobori 2011), this approach allowed us to avoid making assumptions about several variables that have not been quantified at the cell-type level, such as the relative decoding efficiency of wobble and Watson–Crick base pairs and the levels of ANN tRNAs that are modified to INN or left unmodified, which should theoretically scale with the expression of adenosine deaminases across cell types.

We calculated tTE across all mouse and human cell types for which scRNA-seq data were available for AA demand quantification and scATAC-seq data had sufficient resolution for reliable AA supply quantification. We found that tTEs were similar in the mouse (range: $\rho \sim 0.66-0.85$) and in humans (range: $\rho \sim 0.60-0.86$) (Fig. 6A; Supplemental Tables S8, S15). This relatively narrow range of tTEs was consistent with our observations of stability in both codon and anticodon pools across the majority of cell types and was also in line with previous bulk analyses using the same translation efficiency metric (Kutter et al. 2011).

To assess whether these tTEs were statistically significant, we sampled from a null distribution of tTEs by shuffling 1000 times the tRNA gene usage of each cell type, pooling the gene expression levels based on each tRNA's charged AA to calculate AA supply, and then correlating the AA supply to the AA demand from the mRNA side, which was unshuffled (Fig. 6B). After estimating parameters for this null distribution, we compared the values of the observed and simulated tTEs for each cell type. By using this approach, we discovered that all cell types in the mouse and all but one cell type (skeletal muscle cells) in humans are statistically significant (P-value < 0.05) (Fig. 6C). In other words, codon and anticodon pools were not only stable across different mammalian cell types but also strongly correlated with each other, thereby establishing an efficient interface for translation elongation.

Neurons have enhanced translation efficiency

Although most cell types have similar codon and anticodon pools, some cell types were outliers in either codon usage (cardiac/skeletal muscle cells and pancreatic acinar cells) (Fig. 2) or anticodon usage (brain neurons) (Fig. 4) in the mouse and human data sets. We therefore wondered whether the tTE varies in these outliers. We found that cardiac/skeletal muscle cells had among the lowest tTEs and pancreatic acinar cells had moderate tTEs compared with other cell types (Fig. 6A; Supplemental Tables S8, S15). Although cardiac/skeletal muscle cells differed in codon usage largely because of *TTN*'s immense contribution to their codon pool (Fig. 2G), their anticodon pools did not differ. This lack of anticodon compensation for differential codon usage led to the lowest tTE in cardiac/skeletal muscle cells. In contrast, we found that brain neurons have among the highest tTEs. When comparing brain neurons against all other cell types, we observed that they had a statistically significant increase in both tTE and tTE P-value (Fig. 6A,C). Because we determined that codon usage in neuronal cells was comparable to most other cell types, this increased tTE was anticodon driven.

Because we define tTE as the correlation between AA supply and demand, its values are influenced by the similar proportions of each AA's supply and demand. Therefore, we calculated the ratio of AA supply and demand for each of the standard 20 AAs and compared these supply–demand ratios between brain neurons and other cell types. Ala was the only AA with an AA supply-to-demand ratio that had a statistically significant increase in brain neurons compared with other cell types, which remained consistent after splitting leucyl-, seryl-, and arginyl-tRNA pools based on non-wobble positions (Fig. 6D, Supplemental Figs. S8–S9). This is consistent with our finding that Ala anticodons were enriched in brain neurons (Fig. 4). Thus, Ala is a prominent driver of increased tTE in brain neurons.

Increased neuronal supply of the Ala-AGC anticodon results in faster decoding of the Ala-GCC codon

Because the Ala supply–demand ratio was enhanced in neurons, we hypothesized that it could result in faster decoding of Ala codons. The relative decoding rate can be quantified by ribosome profiling, in which ribonuclease-protected RNAs, including ribosome-protected mRNA fragments (RPFs), are sequenced genome-wide (Ingolia et al. 2019). By mapping RPFs to the genome, codons occupying the ribosome E (exit), P (peptidyl-), and A (aminoacyl-) sites can be determined. Because codon–anticodon recognition takes place within the A-site, codons that are slowly decoded by the ribosome are expected to frequently occupy the A-site of RPFs, after adjusting for the background RPF codon frequency. In contrast, faster decoded codons should be less frequent in the A-site of RPFs (Ingolia 2014). Consistent with these expectations, budding yeasts in which specific codon–anticodon interactions were disrupted have significantly higher levels of those impaired codons in RPF A-sites compared with the wild type, suggesting slower decoding (Nedialkova and Leidel 2015).

To test our hypothesis of faster neuronal decoding of Ala codons, we analyzed publicly available cell type–specific ribosome

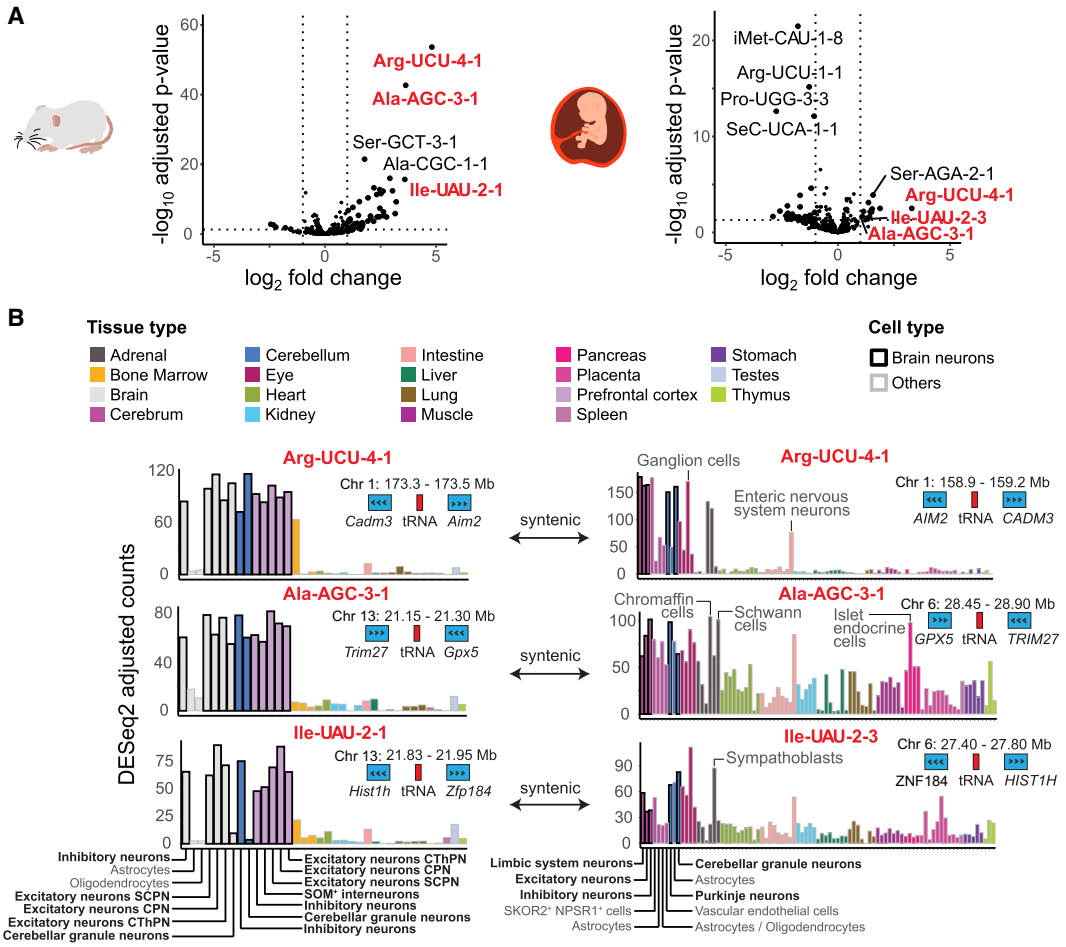


Figure 5. Some tRNA genes are enriched in brain neurons and conserved in the mouse and human. (*A*) Volcano plots show individual tRNA gene usage between brain neurons and all other cell types, displaying $-\log_{10}$ adjusted *P*-values and $\log_2$ FC, as determined using DESeq2. Red labels indicate syntenic tRNA genes significantly enriched in brain neurons in both the mouse (*left*) and human (*right*). Vertical lines indicate a FC >100%. (*B*) Bar plots show expression of syntenic, neuron-enriched tRNA genes. Each bar corresponds to a cell type filled according to the tissue type (*top*) and with border shading for brain neuron (black) and others (gray). Brain cell types (*bottom*), genomic location of the tRNA gene (red) with flanking protein-coding genes (blue) (*top*), and direction of gene transcription (arrows) are indicated.

profiling data in the mouse brain (Scheckel et al. 2020). Because single-cell ribosome profiling technologies have not yet been developed, Scheckel et al. (2020) created mouse cell lines with a GFP-labeled form of a ribosomal protein induced by Cre recombinase under a cell type–specific promoter. Subsequent affinity purification with an anti-GFP antibody was used to obtain RPFs specific to neuron and glial populations (Fig. 6E). After aligning RPFs to the A-sites of protein-coding genes with CONCUR (Frye and Bornelöv 2021), we performed differential analysis between all neuronal and glial samples to examine whether any codons were differentially present in the A-site. Of these codons, Ala-GCC showed the strongest enrichment in glial cells, verifying that neurons decode GCC faster. This is consistent with our analysis of Ala anticodon usage in brain at the single-cell level. We observed that both neurons and glial cells were enriched for Ala-CGC and Ala-UGC compared with nonbrain cell types but were not differentially expressed between glial cells and neurons (Supplemental Fig. S3). In contrast, $tRNA^{Ala}$ (AGC) was specifically enriched in neurons and is the only Ala anticodon that can decode Ala-GCC (Fig. 6F). The interaction between the Ala-GCC codon and the $tRNA^{Ala}$ (AGC) anticodon must occur through a A34-to-I modification, in which the first anticodon position A becomes an I that can then base pair with C, the third codon position. Although cell type–specific ribosome profiling data sets are not yet available in other cell types for which we calculated tTE, these findings underscore that differential codon and anticodon usage can lead to cell type–specific differences in decoding rates during translation elongation.

Moreover, our identification of an increased $tRNA^{Ala}$ (AGC) anticodon pool in neurons may explain a recent finding regarding the role of TRM1L, a tRNA-modification enzyme. Previous studies showed that *Trm1l*-deficient mice had neurological deficits, including altered motor coordination and aberrant exploratory behavior (Vauti et al. 2007). However, the function of this protein was unknown. A recent study described that the TRM1L protein performs *N*2,*N*2-dimethylguanosine ($m^{2,2}G$) modifications at position 26 specifically in $tRNA^{Ala}$ (AGC) (Jonkhout et al. 2021). In addition, TRM1L subcellular localization changes upon neuronal activation but not under general stress, suggesting that this protein plays a role in long-term potential and synaptic plasticity. Although it was hypothesized that the neurological phenotype associated with the *Trm1l* deletion could be caused by increased

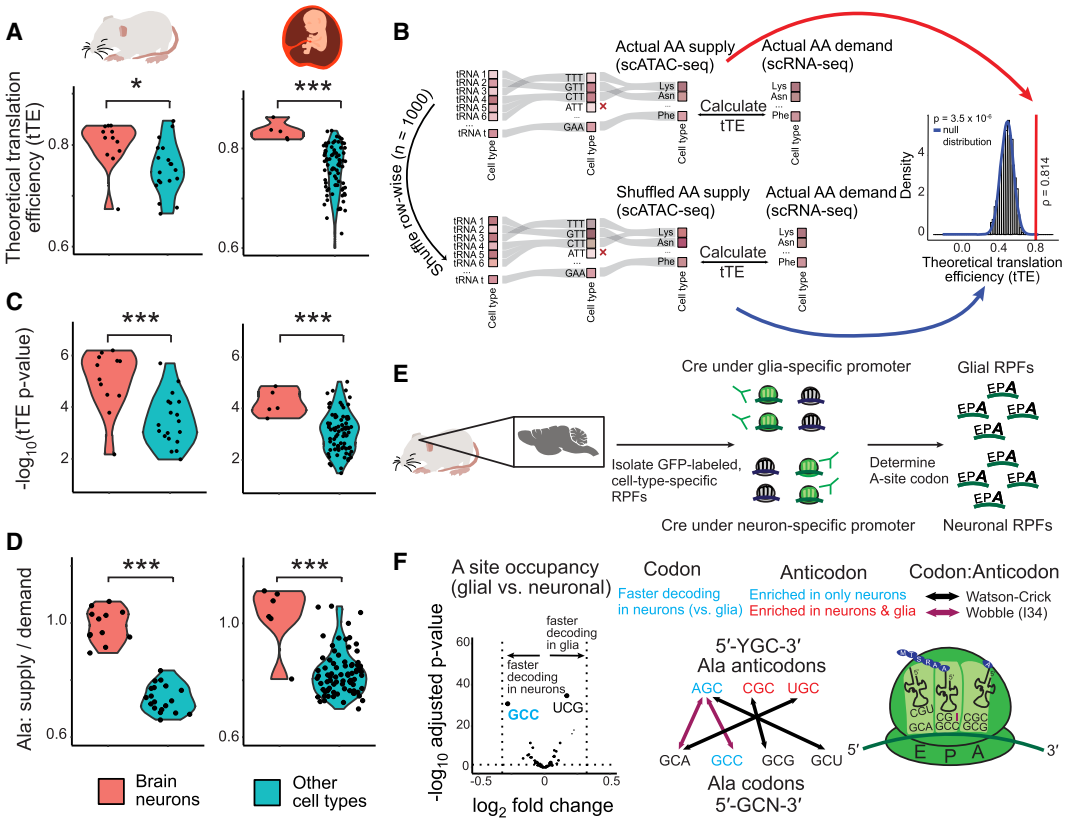


Figure 6. Correlation between AA supply and demand is high and statistically significant in all cell types, with the strongest correlation in brain neurons. (*A*,*C*,*D*) Violin plots are divided into mouse (*left*) and human (*right*) showing tTEs (*A*), tTE *P*-values (*C*), and alanine (Ala) supply to demand ratio (*D*) of brain neurons and other cell types. Of all 20 AAs, only Ala is significantly enriched between brain neurons and all other cell types. (*B*) Schematic representation of the approach used to determine statistical significance of correlation between AA supply (from tRNA side) and AA demand (from mRNA side), defined here as theoretical translation efficiency (tTE). The observed tRNA expression for each cell type is shuffled 1000 times, pooled at the AA supply level, and correlated to the AA demand (unshuffled) to detect a null distribution of tTE values and determine statistical significance of actual tTE. (*E*) Workflow from an adult mouse brain ribosome profiling data set (Scheckel et al. 2020). Cell type–specific ribosome-protected fragments (RPFs) for neurons and glia were obtained with cell type–specific GFP-labeling of a ribosomal protein, followed by immunoprecipitation against GFP. The ribosome A-site of these cell type–specific RPFs was determined, and a differential analysis was performed using DESeq2. (*F*) Volcano plot shows enrichment of $tRNA^{Ala}$ (AGC) anticodon in neurons, and faster decoding of the $tRNA^{Ala}$ (GCC) anticodon is observed in neurons compared with glial cells. AGC must decode GCC via an adenosine to inosine modification at the first anticodon position. Asterisks display degree of significance: (*) $P<0.05$, (**) $P<0.01$, (***) $P<0.001$; Mann–Whitney *U* test.

$tRNA^{Ala}$ (AGC) abundance in the brain, differential expression was masked in bulk sequencing data. But with single-cell sequencing, we detected this neuronal increase in $tRNA^{Ala}$ (AGC) anticodon usage. Modifications influence the stability of tRNA transcripts, and hypomodified tRNA molecules are rapidly degraded (Kimura and Waldor 2019). Thus, absence of the $m^{2,2}G_{26}$ modification upon the deletion of TRM1L may reduce $tRNA^{Ala}$ (AGC) pools in neurons and impact the rate of translation elongation. Future systematic studies may explore whether the absence of tRNA modifications reduces specific anticodon pools and drives malignant phenotypes by slower decoding of the corresponding mRNA codons.

Discussion

Although codon and anticodon usage have been previously quantified in bulk tissue, analysis of these critical players in translation elongation had not yet been explored using single-cell information. In bulk tissue, mixed signals owing to cell heterogeneity were a limiting factor (Schmitt et al. 2014; Rudolph et al. 2016) that was less pronounced in more homogeneous tissues like liver (>70% hepatocytes) but could have a greater importance in tissues with more diverse cell types like the brain. By harnessing publicly available scRNA-seq and scATAC-seq atlases, we have taken these previous analyses one step further and simultaneously analyzed codon usage and AA demand as well as anticodon usage and AA supply across individual cell types from multiple tissues in two mammals (31 in the mouse, 85 in humans). The highly comprehensive nature of these atlases allowed us to examine codon and anticodon usage not only in greater depth but also in breadth because the mouse and human atlases contain more tissues than any bulk data set.

Because existing tRNA quantification methods cannot yet be performed at the single-cell level, we investigated the feasibility of using scATAC-seq and showed its robustness for measuring tRNA gene usage. The increasingly lower cost and proliferation of many fast and reproducible scATAC-seq methods may make this approach especially attractive for tRNA quantification. In particular, scATAC-seq breaks a major bottleneck that has hindered comprehensive analysis of tRNAs across the diverse cell types in

complex mammalian systems. Although our analysis of codon and anticodon usage data may have biases arising from integrating data from two different techniques (scRNA-seq and scATAC-seq) performed on different samples, the increased proliferation of multiomic techniques that can jointly quantify chromatin accessibility and transcripts in single cells should remove this limitation (Swanson et al. 2021). Another issue with scATAC-seq is that few cuts to tRNA genes are sequenced under normal sequence depths, which necessitated removing some cell types with insufficient cuts from our analysis (Fig. 3C). However, a multiomic approach that combines scRNA-seq (for cell type annotation) and single-cell CUT&Tag (Kaya-Okur et al. 2019) to enrich specifically for Tn5 insertions near Pol III binding sites should mitigate this problem.

Although we identified several features of codon usage, anticodon usage, and translation efficiency that are conserved over 90 million years of mammalian evolution (Kumar et al. 2017), we found an increase of Ala supply in the mouse and human brain across all data sets (bulk and single cell). Besides Ala, other AAs are enriched in some brain data sets but not in others. As a result, the apparent faster decoding of the codon Cys-UCG in glial cells within the ribosome profiling data set remains to be resolved (Fig. 6F). AA supply and demand are known to change at timescales of minutes and hours. These fluctuations across data sets may reflect differences right before sample collection (Rak et al. 2018). In contrast, the evolutionarily conserved nature of increased $tRNA^{Ala}$ (AGC) supply in both mammals at different time points suggests purifying selection and a longer-term significance in organismal development.

Although Ala supply is increased in neurons, we did not see corresponding changes in anticodon usage to meet codon demand for skeletal and cardiac muscle cells. This imbalance is largely driven by the codon usage of a single gene, *Ttn*. When this gene is removed from the tTE calculation, skeletal and cardiac muscle cells have a similar tTE to most other cell types. Whether post-transcriptional regulation resolves the potential translational choke point caused by imbalance of codon demand and anticodon supply for *Ttn* is unknown and will require further experimental investigation.

Another potential area for progress involves better modeling of the rate of translation elongation. When calculating translation efficiency, we determined AA supply from the usage of all tRNA genes charged with the same AA and calculated AA demand by weighting the AA gene usages by their expression for all cell types. From the AA supply side, it is oversimplistic to assume that all tRNAs will necessarily function in translation elongation, especially because tRNAs are known to perform an array of different roles, including as tRNA-derived fragments (Polacek and Ivanov 2020). From the AA demand side, different protein isoforms are known to exist across cell types. We did not account for this owing to the lack of data availability, although the stability of codon usage across cell types that express entirely different proteins may indicate that including isoform data will not affect our main conclusions. For both AA supply and demand, it is important to consider not only the abundances of tRNAs and mRNAs but also their turnover rates. Once single-cell half-life information of mRNA and tRNA species becomes feasible to measure, AA supply and demand should account not only for expression but also for the differential stability of these transcripts. Additionally, codon optimality has been shown to be a major determinant of mRNA stability (Presnyak et al. 2015). Thus, the ratios of corresponding codons and anticodons or AA supply and demand across different cell types could be correlated with the stability of mRNA transcripts. In particular, it would be interesting to verify if neuronal depletion of the $tRNA^{Ala}$ (AGC) anticodon results in slower decoding of Ala-GCC codons to induce slowness-mediated decay of certain transcripts vital for neuronal function (Rak et al. 2018).

For translation efficiency, we calculated tTE at the level of AA supply and demand, rather than at the codon–anticodon level, as it is unclear whether and how to consider wobble base-pairing and other modulators of the codon–anticodon interaction such as tRNA modification enzymes. Moreover, there may be tissue-specific tRNA modifications, including a recent finding that some $tRNA^{Ala}$ anticodons are enriched for particular modifications in the brain (Pinkard et al. 2020). Thus, another avenue of investigation will involve disentangling cell type differences in these processes that could impact the rate of translation elongation.

Finally, it is worth noting that our analyses were performed on atlases of mice and human samples believed to represent healthy states. Thus, the stability that we observe in codon and anticodon pools may exist only in a healthy state, and dysregulation of these pools may occur in abnormal states such as cancer (Goodarzi et al. 2016; Zhang et al. 2018), in neurodegenerative diseases (Kapur et al. 2017), and with disruption to the microbiome (Schwartz et al. 2018; Huang et al. 2021). Although such single-cell data are not yet available, approaches similar to those presented here could be used to examine cell type–specific changes to codon usage, anticodon usage, and translation efficiency in disease.

Methods

Data sets analyzed

All data analyzed were downloaded from publicly available sources (for more information, see Supplemental Table S1). The mouse scRNA-seq atlas was obtained from the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE109774 (The Tabula Muris Consortium 2018), and Seurat objects were downloaded from the *Tabula Muris* website (https://tabula-muris.ds.czbiohub.org/). The mouse scATAC-seq atlas (Cusanovich et al. 2018) was obtained from GEO accession number GSE111586, and BAM files were downloaded from the website (https://atlas.gs.washington.edu/mouse-atac/). The human scRNA-seq (Cao et al. 2020) and scATAC-seq (Domcke et al. 2020) atlases were retrieved from GEO accession numbers GSE156793 and GSE149683, respectively. Count matrices (scRNA-seq) and fragment files (scATAC-seq) were downloaded from the website (https:// descartes.brotmanbaty.org/bbi/). Two other mouse brain scATAC-seq data sets were analyzed to determine scATAC-seq reproducibility (Fig. 3C; Supplemental Fig. S2): a 10x Genomics data set (atac v1 adult brain fresh 5k) and a droplet-based data set from Lareau et al. (2019) (GEO accession number GSE123581). scATAC-seq tRNA gene usage was also compared with bulk adult mouse ATAC-seq data (NCBI Sequence Read Archive [SRA; https://www.ncbi.nlm.nih.gov/sra] accession number SRX4946150) (Liu et al. 2019) and ChIP-seq data (ArrayExpress [https://www.ebi.ac.uk/arrayexpress/] accession number E-MTAB-2326) (Schmitt et al. 2014). To compare with anticodon usage results from scATAC-seq concerning increased alanine supply, bulk experiments were analyzed. A bulk ATAC-seq atlas of early mouse development was downloaded from https://www.encodeproject.org/ (Gorkin et al. 2020). Bulk ChIP-seq data of mouse liver and brain across development were downloaded from ArrayExpress accession E-MTAB-2326 (Schmitt et al. 2014). QuantM-tRNA-seq of adult mouse tissues was downloaded from GEO accession number GSE141436 (Pinkard et al. 2020). To examine differential decoding of codons in neurons versus glial cells,

the cell type–specific mouse brain ribosomal profiling data set was downloaded from GEO accession number GSE149805 (Scheckel et al. 2020). For consistency with former studies and reusage of the intermediate files provided, we used the same reference genome for the mouse (*Mus musculus*, GRCm38) and humans (*Homo sapiens*, GRCh37). Using GRCh37 instead of the more recent GRCh38 would not significantly affect our conclusions because results are highly reproducible with both genome versions.

Data analysis and visualization were conducted in R (R Core Team 2021). Tidyverse was used for data exploration and analysis, and ggplot was used for data visualization. Several packages in the Bioconductor suite were used (Gentleman et al. 2004). Seurat was used for scRNA-seq analysis (Satija et al. 2015), and Signac was used for scATAC-seq analysis (Stuart et al. 2021). DESeq2 was used for differential gene analysis (Love et al. 2014). tRNAscanImport (https://github.com/FelixErnst/tRNAscanImport) was used to load tRNA gene predictions for GRCm38 and GRCh37. Gviz was used for genome browser visualization (Hahne and Ivanek 2016).

Count matrices, pooled at the cell-type level, for mRNA gene expression, codon usage, and AA demand, as well as for tRNA gene expression, anticodon usage, and AA supply, are available as Supplemental Data (Supplemental Tables S2–S7 and S9–S14 for mouse and human, respectively).

Quantification of codon usage and AA demand from scRNA-seq data sets

Codon usage for each protein-coding gene was determined using the Ensembl set of protein-coding genes for mouse (GRCm38) and human (GRCh37). This was then formatted as a codon frequency per gene matrix of dimensions 61 × p, where the rows correspond to the 61 sense codons and the columns to each of the p protein-coding genes. Codon usage for each individual cell was determined by matrix-multiplying the codon frequency per gene matrix by the scRNA-seq count matrix, formatted as a $p \times n$ matrix, where p again corresponds to the protein-coding genes (ordered in the same way as the codon usage matrix) and n to the number of cells remaining after quality control filtering. Only genes present in both the count matrix and the codon usage were used, because p must be the same for matrix multiplication. The multiplication of these two matrices weights the codon frequencies of each protein-coding gene by the expression in each cell, producing a "codon usage per cell matrix" of dimension 61 × n.

Using cell type annotations provided by the atlases used in this study, we pooled the codon usage of cells of the same cell type, producing a "codon usage per cell type" matrix of dimension 61 × m, where m denotes the number of cell types annotated in the experiment. Each codon has a specific AA demanded per standard genetic code. Therefore, AA demanded can be quantified by pooling codons demanding the same AA to produce an "AA per cell type" matrix of dimension 20 × m, where the rows correspond to the 20 classical AAs.

Quantification of tRNA gene usage, anticodon usage, and AA supply from scATAC-seq data sets

tRNA gene annotations were used from gtRNAdb (Chan and Lowe 2016) that uses tRNAscan-SE v2.0 to predict tRNA genes. tRNAscan-SE denotes functional tRNAs as "high confident" based on predicted secondary structure stability and other tRNA-specific features, whereas "low confidents" are considered as probable pseudogenes. There are 401 high-confidence tRNA genes in the adult mouse (GRCm38 annotation) and 416 in humans (GRCh37 annotation). Quantification of tRNA gene usage from scATAC-seq data sets required generating a tRNA count matrix of dimension $t \times n$, where t corresponds to the number of high-confidence tRNA genes. For scATAC-seq data sets, this quantification was performed using the FeatureMatrix function available in Signac, which uses a fragment file and a set of genomic regions as input (Stuart et al. 2021). Fragment files were already available for the human data set. For the mouse data set, BAM files were downloaded from https://atlas.gs.washington.edu/mouse-atac/, and sinto (https://timoast.github.io/sinto/) was used for generation of fragment files. FeatureMatrix quantifies the number of Tn5 insertions occurring within each of the specified genomic regions (formatted as a Granges object). In our case, the genomic regions were the set of high-confidence tRNA genes, including the gene body and 100 bp upstream and downstream, as has been performed in many other studies using Pol III ChIP-seq (Kutter et al. 2011).

This "tRNA gene expression per cell" matrix can be pooled by cell type annotation to produce a "tRNA gene expression per cell type" matrix of dimension $t \times m$, where m refers to the number of cell types annotated. This matrix was then used to examine the total number of cuts for each tRNA gene across all cells (pseudobulked), and a bimodal distribution was observed, interpreted as background and true signal distribution. tRNA genes belonging to the background distribution were removed from the analysis. tRNA genes can then be pooled based on their anticodon sequences to produce an "anticodon usage per cell type" matrix of dimension $a \times m$, where a refers the number of unique anticodons (46 in GRCm38 and 47 in GRCh37). Finally, AA supply can be quantified by pooling anticodon families by the AA they accept, resulting in an "AA supply per cell type" matrix of dimension 21 × m. The 21st AA corresponds to selenocysteine.

Differential analysis

Differential tRNA gene expression, anticodon usage, and AA supply analysis was performed using DESeq2 (Love et al. 2014) under default settings, with input being the tRNA gene, anticodon usage, and AA supply per cell type matrices, respectively. Brain neurons were compared against all other cell types in the mouse and human data sets.

Translation efficiency analysis

We computed tTE as the Spearman's rank correlation coefficient between the AA demand from the mRNA codon side and the AA supply from the tRNA anticodon side. In other words, the values for each cell type in the "AA demand per cell type" matrix and the "AA supply per cell type" matrix were correlated with each other for all cell types that were present in both the scRNA-seq and scATAC-seq data sets. For this analysis, selenocysteine was ignored because the number of stop codons that bind to the selenocysteine tRNA is unknown at the cell-type level.

Ribosome profiling analysis

Scheckel et al. (2020) performed ribosome profiling on mice injected with control homogenate or prions and extracted cell type–specific ribosome-protected fragments pertaining to CamKIIa excitatory neurons, PV interneurons, microglia, and astrocytes. They collected samples from mice sacrificed 2, 4-, 8-, 16-, and 24-wk post injection, as well as samples collected at a later point before severe prion disease developed in the prion-infected group. We downloaded the ribosome profiling data for all time points for the control groups. Fewer RPFs at the A-site, when corrected for background codon frequencies at other positions in RPFs, suggests faster decoding of a particular codon. To identify A-site codon for

each RPF, we used CONCUR (Frye and Bornelöv 2021). CONCUR accounts for potential differences in the codon usages of the translated mRNA populations between different cell types by dividing the codon frequencies in the A-site by the codon frequencies of all codons within the RPF. RPFs are usually about nine to 10 codons long for typical experimental protocols, and so, determining the codon frequency from the entire RPF gives a background frequency of translated codons in that cell type. To identify codons that were differentially found in the ribosome A-site, we used DESeq2 (Love et al. 2014), as described above, to compare the neuronal group (CamKIIa excitatory neuron and PV interneuron samples at all time points) and the glial group (microglia and astrocyte cell types), while controlling for injection time.

Software availability

The R code, and detailed instructions on how to preprocess the data and reproduce the results, are available as Supplemental Code and at GitHub (https://github.com/wgao688/sc_tRNA_mRNA).

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank Konrad Rudolph for critical comments on the manuscript. We thank Keyi Geng and Marcel Tarbier as well as other group members of the laboratories of Claudia Kutter, Marc Friedländer, and Vicent Pelechano for helpful feedback regarding the experimental procedures, data analysis, and data presentation. We thank the Tabula Muris and Descartes consortia for sharing invaluable data to the research community. This work was supported by the U.S.–Sweden Fulbright Student Research Program (W.G.), Knut & Alice Wallenberg Foundation (KAW 2016.0174; C.K.), Ruth & Richard Julin Foundation (2017–00358, 2018–00328, 2020-00294; C.K.), SFO-SciLifeLab fellowship (SFO_004; C.K.), Swedish Research Council (2019–05165; C.K.), Lillian Sagen & Curt Ericsson Research Foundation (2021-00427; C.K.), Gösta Miltons Research Foundation (2021-00527, C.K.), and the Swedish National Infrastructure for Computing projects at UPPMAX (storage: 2021/22-133; compute: 2021/22-133).

Author contributions: W.G. and C.K. conceptualized the project. W.G., C.J.G-D., and C.K. designed the analyses. W.G. performed the analyses. W.G., C.J.G-D., and C.K. wrote the original draft.

References

Baek S, Lee I. 2020. Single-cell ATAC sequencing analysis: from data preprocessing to hypothesis generation. *Comput Struct Biotechnol J* **18:** 1429–1439. doi:10.1016/j.csbj.2020.06.012

Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. 2015. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol* **109:** 21.29.1–21.29.9. doi:10.1002/0471142727.mb2129s109

Cao J, O'Day DR, Pliner HA, Kingsley PD, Deng M, Daza RM, Zager MA, Aldinger KA, Blecher-Gonen R, Zhang F, et al. 2020. A human cell atlas of fetal gene expression. *Science* **370**. eaba7721. doi:10.1126/science.aba7721

Chan PP, Lowe TM. 2016. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res* **44:** D184–D189. doi:10.1093/nar/gkv1309

Chou CC, Chen CH, Lee TT, Peck K. 2004. Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression. *Nucleic Acids Res* **32:** e99. doi:10.1093/nar/gnh099

Cusanovich DA, Hill AJ, Aghamirzaie D, Daza RM, Pliner HA, Berletch JB, Filippova GN, Huang X, Christiansen L, DeWitt WS, et al. 2018. A single-cell atlas of *in vivo* mammalian chromatin accessibility. *Cell* **174:** 1309–1324.e18. doi:10.1016/j.cell.2018.06.052

Dever TE, Dinman JD, Green R. 2018. Translation elongation and recoding in eukaryotes. *Cold Spring Harb Perspect Biol* **10:** a032649. doi:10.1101/cshperspect.a032649

Dieci G, Fiorino G, Castelnuovo M, Teichmann M, Pagano A. 2007. The expanding RNA polymerase III transcriptome. *Trends Genet* **23:** 614–622. doi:10.1016/j.tig.2007.09.001

Dittmar KA, Goodenbour JM, Pan T. 2006. Tissue-specific differences in human transfer RNA expression. *PLoS Genet* **2:** e221. doi:10.1371/journal.pgen.0020221

Domcke S, Hill AJ, Daza RM, Cao J, O'Day DR, Pliner HA, Aldinger KA, Pokholok D, Zhang F, Milbank JH, et al. 2020. A human cell atlas of fetal chromatin accessibility. *Science* **370**. eaba7612. doi:10.1126/science.aba7612

dos Reis M, Wernisch L. 2009. Estimating translational selection in eukaryotic genomes. *Mol Biol Evol* **26:** 451–461. doi:10.1093/molbev/msn272

dos Reis M, Savva R, Wernisch L. 2004. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res* **32:** 5036–5044. doi:10.1093/nar/gkh834

Frye M, Bornelöv S. 2021. CONCUR: quick and robust calculation of codon usage from ribosome profiling data. *Bioinformatics* **37:** 717–719. doi:10.1093/bioinformatics/btaa733

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5:** R80. doi:10.1186/gb-2004-5-10-r80

Gogakos T, Brown M, Garzia A, Meyer C, Hafner M, Tuschl T. 2017. Characterizing expression and processing of precursor and mature human tRNAs by hydro-tRNAseq and PAR-CLIP. *Cell Rep* **20:** 1463–1475. doi:10.1016/j.celrep.2017.07.029

Goodarzi H, Nguyen HCB, Zhang S, Dill BD, Molina H, Tavazoie SF. 2016. Modulated expression of specific tRNAs drives gene expression and cancer progression. *Cell* **165:** 1416–1427. doi:10.1016/j.cell.2016.05.046

Gorkin DU, Barozzi I, Zhao Y, Zhang Y, Huang H, Lee AY, Li B, Chiou J, Wildberg A, Ding B, et al. 2020. An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature* **583:** 744–751. doi:10.1038/s41586-020-2093-3

Hahne F, Ivanek R. 2016. Visualizing genomic data using Gviz and Bioconductor. *Methods Mol Biol* **1418:** 335–351. doi:10.1007/978-1-4939-3578-9_16

Huang J, Chen W, Zhou F, Pang Z, Wang L, Pan T, Wang X. 2021. Tissue-specific reprogramming of host tRNA transcriptome by the microbiome. *Genome Res*, **31:** 947–957. doi:10.1101/gr.272153.120

Ingolia NT. 2014. Ribosome profiling: new views of translation, from single codons to genome scale. *Nat Rev Genet* **15:** 205–213. doi:10.1038/nrg3645

Ingolia NT, Hussmann JA, Weissman JS. 2019. Ribosome profiling: global views of translation. *Cold Spring Harb Perspect Biol* **11:** a032698. doi:10.1101/cshperspect.a032698

Ishimura R, Nagy G, Dotu I, Zhou H, Yang XL, Schimmel P, Senju S, Nishimura Y, Chuang JH, Ackerman SL. 2014. Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. *Science* **345:** 455–459. doi:10.1126/science.1249749

Jonkhout N, Cruciani S, Santos Vieira HG, Tran J, Liu H, Liu G, Pickford R, Kaczorowski D, Franco GR, Vauti F, et al. 2021. Subcellular relocalization and nuclear redistribution of the RNA methyltransferases TRMT1 and TRMT1L upon neuronal activation. *RNA Biol* **18:** 1905–1919. doi:10.1080/15476286.2021.1881291

Kapur M, Monaghan CE, Ackerman SL. 2017. Regulation of mRNA translation in neurons: a matter of life and death. *Neuron* **96:** 616–637. doi:10.1016/j.neuron.2017.09.057

Kapur M, Ganguly A, Nagy G, Adamson SI, Chuang JH, Frankel WN, Ackerman SL. 2020. Expression of the neuronal tRNA n-Tr20 regulates synaptic transmission and seizure susceptibility. *Neuron* **108:** 193–208.e9. doi:10.1016/j.neuron.2020.07.023

Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K, Henikoff S. 2019. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10:** 1930. doi:10.1038/s41467-019-09982-5

Kimura S, Waldor MK. 2019. The RNA degradosome promotes tRNA quality control through clearance of hypomodified tRNA. *Proc Natl Acad Sci* **116:** 1394–1403. doi:10.1073/pnas.1814130116

Klemm SL, Shipony Z, Greenleaf WJ. 2019. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20:** 207–220. doi:10.1038/s41576-018-0089-8

Kugelberg U, Nätt D, Skog S, Kutter C, Öst A. 2021. 5´XP sRNA-seq: efficient identification of transcripts with and without 5´ phosphorylation reveals evolutionary conserved small RNA. *RNA Biol* **18:** 1588–1599. doi:10.1080/15476286.2020.1861770

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol* **34:** 1812–1819. doi:10.1093/molbev/msx116

Kutter C, Brown GD, Gonçalves Å, Wilson MD, Watt S, Brazma A, White RJ, Odom DT. 2011. Pol III binding in six mammals shows conservation among amino acid isotypes despite divergence among tRNA genes. *Nat Genet* **43:** 948–955. doi:10.1038/ng.906

Lareau CA, Duarte FM, Chew JG, Kartha VK, Burkett ZD, Kohlway AS, Pokholok D, Aryee MJ, Steemers FJ, Lebofsky R, et al. 2019. Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat Biotechnol* **37:** 916–924. doi:10.1038/s41587-019-0147-6

Lewinter MM, Granzier H. 2010. Cardiac titin: a multifunctional giant. *Circulation* **121:** 2137–2145. doi:10.1161/CIRCULATIONAHA.109.860171

Liu C, Wang M, Wei X, Wu L, Xu J, Dai X, Xia J, Cheng M, Yuan Y, Zhang P, et al. 2019. An ATAC-seq atlas of chromatin accessibility in mouse tissues. *Sci Data* **6:** 65. doi:10.1038/s41597-019-0071-0

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550. doi:10.1186/s13059-014-0550-8

Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* **44:** W54–W57. doi:10.1093/nar/gkw413

Nedialkova DD, Leidel SA. 2015. Optimization of codon translation rates via tRNA modifications maintains proteome integrity. *Cell* **161:** 1606–1618. doi:10.1016/j.cell.2015.05.022

Novoa EM, Ribas de Pouplana L. 2012. Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet* **28:** 574–581. doi:10.1016/j.tig.2012.07.006

Pan T. 2018. Modifications and functional genomics of human transfer RNA. *Cell Res* **28:** 395–404. doi:10.1038/s41422-018-0013-y

Phizicky EM, Hopper AK. 2010. tRNA biology charges to the front. *Genes Dev* **24:** 1832–1860. doi:10.1101/gad.1956510

Pinkard O, McFarland S, Sweet T, Coller J. 2020. Quantitative tRNA-sequencing uncovers metazoan tissue-specific tRNA regulation. *Nat Commun* **11:** 4104. doi:10.1038/s41467-020-17879-x

Polacek N, Ivanov P. 2020. The regulatory world of tRNA fragments beyond canonical tRNA biology. *RNA Biol* **17:** 1057–1059. doi:10.1080/15476286.2020.1785196

Pott S, Lieb JD. 2015. Single-cell ATAC-seq: strength in numbers. *Genome Biol* **16:** 172. doi:10.1186/s13059-015-0737-7

Pouyet F, Mouchiroud D, Duret L, Sémon M. 2017. Recombination, meiotic expression and human codon usage. *eLife* **6:** e27344. doi:10.7554/eLife.27344

Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker KE, Graveley BR, et al. 2015. Codon optimality is a major determinant of mRNA stability. *Cell* **160:** 1111–1124. doi:10.1016/j.cell.2015.02.029

Quax TEF, Claassens NJ, Söll D, van der Oost J. 2015. Codon bias as a means to fine-tune gene expression. *Mol Cell* **59:** 149–161. doi:10.1016/j.molcel.2015.05.035

Rak R, Dahan O, Pilpel Y. 2018. Repertoires of tRNAs: the couplers of genomics and proteomics. *Annu Rev Cell Dev Biol* **34:** 239–264. doi:10.1146/annurev-cellbio-100617-062754

R Core Team. 2021. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.

Rocha EPC. 2004. Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Res* **14:** 2279–2286. doi:10.1101/gr.2896904

Rudolph KLM, Schmitt BM, Villar D, White RJ, Marioni JC, Kutter C, Odom DT. 2016. Codon-driven translational efficiency is stable across diverse mammalian cell states. *PLoS Genet* **12:** e1006024. doi:10.1371/journal.pgen.1006024

Satija R, Farrell JA, Gennert D, Schier AF, Regev A. 2015. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33:** 495–502. doi:10.1038/nbt.3192

Scheckel C, Imeri M, Schwarz P, Aguzzi A. 2020. Ribosomal profiling during prion disease uncovers progressive translational derangement in glia but not in neurons. *eLife* **9:** e62911. doi:10.7554/eLife.62911

Schmitt BM, Rudolph KLM, Karagianni P, Fonseca NA, White RJ, Talianidis I, Odom DT, Marioni JC, Kutter C. 2014. High-resolution mapping of transcriptional dynamics across tissue development reveals a stable mRNA–tRNA interface. *Genome Res* **24:** 1797–1807. doi:10.1101/gr.176784.114

Schwartz MH, Wang H, Pan JN, Clark WC, Cui S, Eckwahl MJ, Pan DW, Parisien M, Owens SM, Cheng BL, et al. 2018. Microbiome characterization by high-throughput transfer RNA sequencing and modification analysis. *Nat Commun* **9:** 5353. doi:10.1038/s41467-018-07675-z

Shigematsu M, Honda S, Loher P, Telonis AG, Rigoutsos I, Kirino Y. 2017. YAMAT-seq: an efficient method for high-throughput sequencing of mature transfer RNAs. *Nucleic Acids Res* **45:** e70. doi:10.1093/nar/gkx005

Stark R, Grzelak M, Hadfield J. 2019. RNA sequencing: the teenage years. *Nat Rev Genet* **20:** 631–656. doi:10.1038/s41576-019-0150-2

Stuart T, Srivastava A, Madad S, Lareau CA, Satija R. 2021. Single-cell chromatin state analysis with Signac. *Nat Methods* **18:** 1333–1341. doi:10.1038/s41592-021-01282-5

Suzuki T. 2021. The expanding world of tRNA modifications and their disease relevance. *Nat Rev Mol Cell Biol* **22:** 375–392. doi:10.1038/s41580-021-00342-0

Swanson E, Lord C, Reading J, Heubeck AT, Genge PC, Thomson Z, Weiss MDA, Li XJ, Savage AK, Green RR, et al. 2021. Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq. *eLife* **10:** e63632. doi:10.7554/eLife.63632

The Tabula Muris Consortium. 2018. Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*. *Nature* **562:** 367–372. doi:10.1038/s41586-018-0590-4

Torres AG, Piñeyro D, Filonava L, Stracker TH, Batlle E, Ribas De Pouplana L. 2014. A-to-I editing on tRNAs: biochemical, biological and evolutionary implications. *FEBS Lett* **588:** 4279–4286. doi:10.1016/j.febslet.2014.09.025

Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, et al. 2015. Tissue-based map of the human proteome. *Science* **347:** 1260419. doi:10.1126/science.1260419

Vauti F, Goller T, Beine R, Becker L, Klopstock T, Hölter SM, Wurst W, Fuchs H, Gailus-Durner V, de Angelis MH, et al. 2007. The mouse *Trm1-like* gene is expressed in neural tissues and plays a role in motor coordination and exploratory behaviour. *Gene* **389:** 174–185. doi:10.1016/j.gene.2006.11.004

Watanabe K, Yokobori SI. 2011. tRNA modification and genetic code variations in animal mitochondria. *J Nucleic Acids* **2011:** 623095. doi:10.4061/2011/623095

White RJ. 2011. Transcription by RNA polymerase III: more complex than we thought. *Nat Rev Genet* **12:** 459–463. doi:10.1038/nrg3001

Wolin SL, Matera AG. 1999. The trials and travels of tRNA. *Genes Dev* **13:** 1–10. doi:10.1101/gad.13.1.1

Wong PBY, Wiley EO, Johnson WE, Ryder OA, O'Brien SJ, Haussler D, Koepfli KP, Houck ML, Perelman P, Mastromonaco G, et al. 2012. Tissue sampling methods and standards for vertebrate genomics. *Gigascience* **1:** 8. doi:10.1186/2047-217X-1-8

Xu H, Yao J, Wu DC, Lambowitz AM. 2019. Improved TGIRT-seq methods for comprehensive transcriptome profiling with decreased adapter dimer formation and bias correction. *Sci Rep* **9:** 7953. doi:10.1038/s41598-019-44457-z

Zhang Z, Ye Y, Gong J, Ruan H, Liu CJ, Xiang Y, Cai C, Guo AY, Ling J, Diao L, et al. 2018. Global analysis of tRNA and translation factor expression reveals a dynamic landscape of translational regulation in human cancers. *Commun Biol* **1:** 234. doi:10.1038/s42003-018-0239-8

Zheng G, Qin Y, Clark WC, Dai Q, Yi C, He C, Lambowitz AM, Pan T. 2015. Efficient and quantitative high-throughput tRNA sequencing. *Nat Methods* **12:** 835–837. doi:10.1038/nmeth.3478

Received June 28, 2021; accepted in revised form November 24, 2021.